Real-time Diagnosis of Anomalies in Deep Space Network Operations using Machine Learning

Umaa Rebbapragada

Jet Propulsion Laboratory,
California Institute of Technology

Jet Propulsion Laboratory
California Institute of Technology

Acknowledgements / Contributions

- Rishi Verma[1]
- Kyongsik (KS) Yun
- John Mason
- James Montgomery
- Lauren Klein (USC), summer intern 2020

[1] Mission-Critical, Real-Time Fault-Detection for NASA's Deep Space Network using Apache Kafka
https://videos.confluent.io/watch/xRmEDYjAhiVu56xcmMy6Yr

Goldstone Complex

Madrid Complex

Canberra Complex

Network Control (JPL)

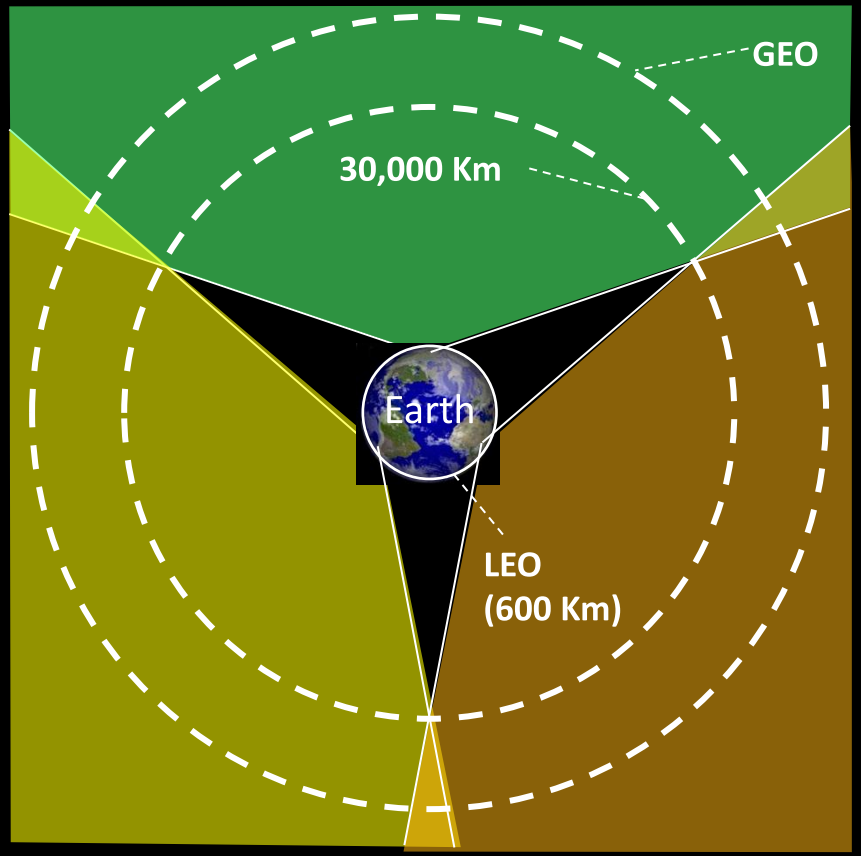
“NASA’s Deep Space Network (DSN) was established in December 1963 to provide a communications infrastructure for deep space missions”

Goldstone Complex

Madrid Complex

Canberra Complex

A Global Enterprise by Necessity



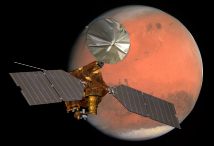
A Diverse Set of Missions

Not all supported spacecraft depicted

Deep Space Communications

Science
- Radar
- Radio Science
- Radio Astronomy

Spacecraft Uplink
- Pointing accuracy
- Antenna efficiency
- Scheduling

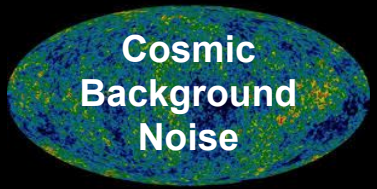


Spacecraft

Hot Body Noise

Spacecraft Downlink
- Signal dispersion
- Noise
- Data rates
- Low power
- Weather
- RFI

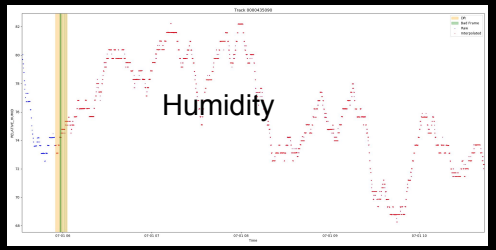
DSN Antenna

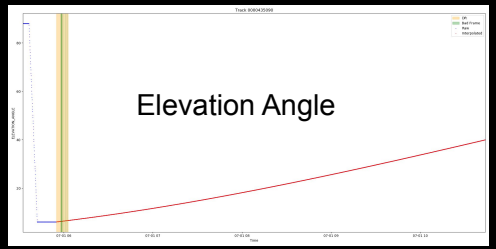


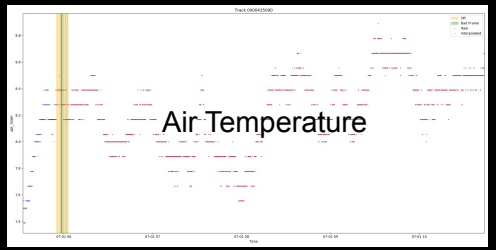
Cosmic Background Noise

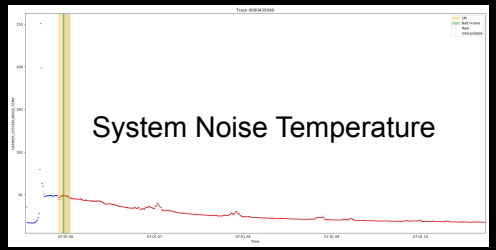
Terminology: Track

- **From a data science perspective, a track is associated with a collection of time series from the beginning to end of DSN communication with the spacecraft that measure:**
  - **Attributes and strength of the signal**
  - **Attributes and telemetry associated with the spacecraft and DSN equipment**
  - **Weather conditions at the site**









Multivariate Time Series

A single track has many time series associated with it, collectively referred to as monitor data

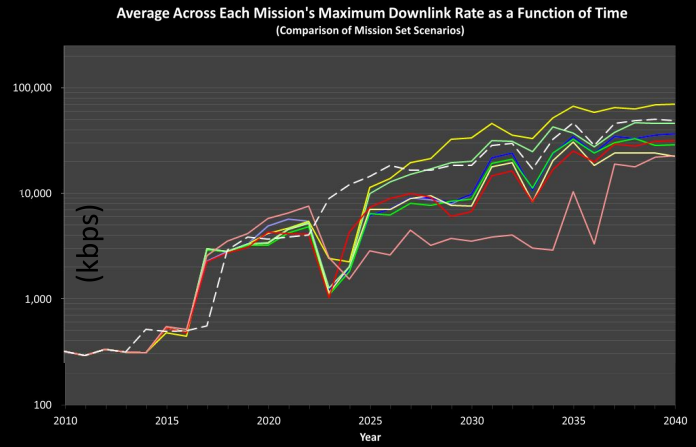
There are hundreds of monitor data items

Challenges

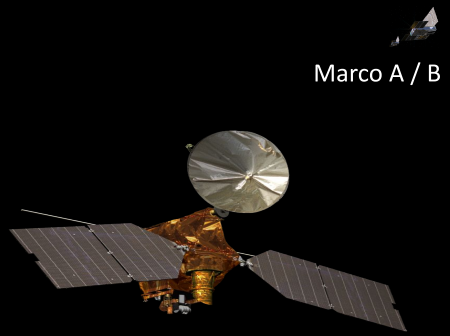
- DSN has pressure to reduce costs while maintaining quality of support to DSN mission users
- DSN complexes were once staffed 24/7 and each operator monitored a single track
- The Follow the Sun initiative (launched 2017) staffs a single complex during their daylight hours only to monitor tracks at all three complexes simultaneously

Operators supporting up to 4 simultaneously tracks

DSN has seen data rates increase over time



More missions, especially cubesats at lunar distances and beyond



Marco A / B

Mars Reconnaissance Orbiter

Matching ML Capabilities with DSN Challenges

- Supervised learning
  - Classification
  - Regression

- Unsupervised learning
  - Clustering
  - Density Estimation
  - Anomaly Detection

Matching ML Capabilities with DSN Challenges

- Supervised learning
  - Classification
  - Regression

- Unsupervised learning ✓
  - Clustering
  - Density Estimation
  - Anomaly Detection

We have historical data stored in a database at a slightly coarser sampling rate.

Matching ML Capabilities with DSN Challenges

- Supervised learning ✓
  - Classification
  - Regression

DSN has a system of reporting problems. Discrepancy Reports can provide some labels for classification. The data also contains actual measurements that can be compared to predicted ones. Thus, regression is possible.

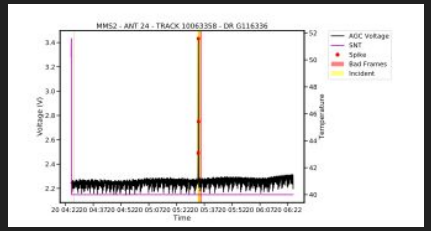
- Unsupervised learning ✓
  - Clustering
  - Density Estimation
  - Anomaly Detection

We have historical data stored in a database at a slightly coarser sampling rate.

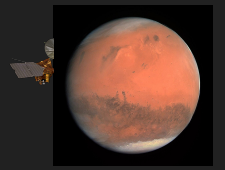
Developing Use Cases

- Team of mostly data scientists and developers
- DSN operations engineer (who formerly worked as an operator) assisting
- Developed four use cases based on feasibility of ML solution
  - supported by initial prototyping
- Surveyed DSN link control operators and operations engineers

1. Compare tracks across all monitor data items for similarity (multivariate time series similarity)
2. Identify anomalies within monitor data items given previous history



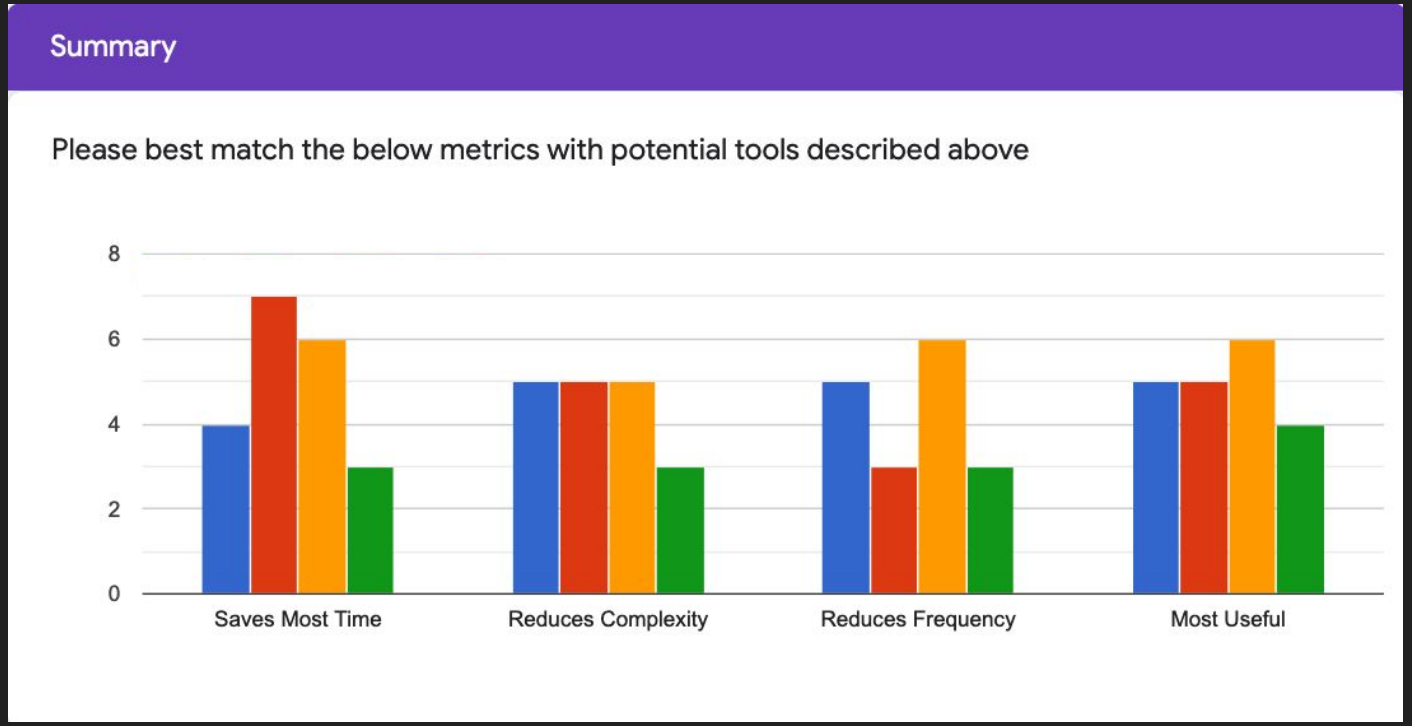
3. Classify whether a track is losing lock with the spacecraft due to bad weather and radio frequency interference (RFI)
4. Detect timing abnormalities in acquisition/loss of spacecraft signal



Survey Results

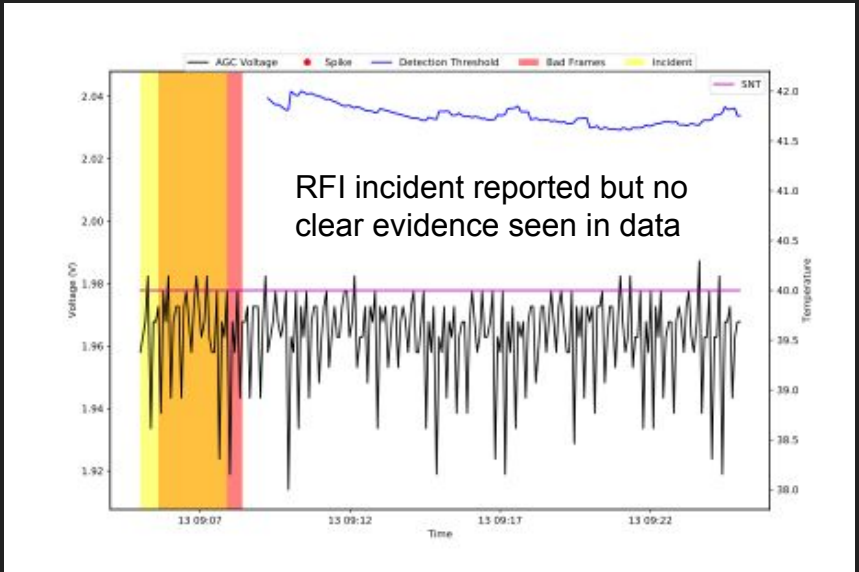
Based off of responses from 21 DSN staff from Madrid, Canberra, and California

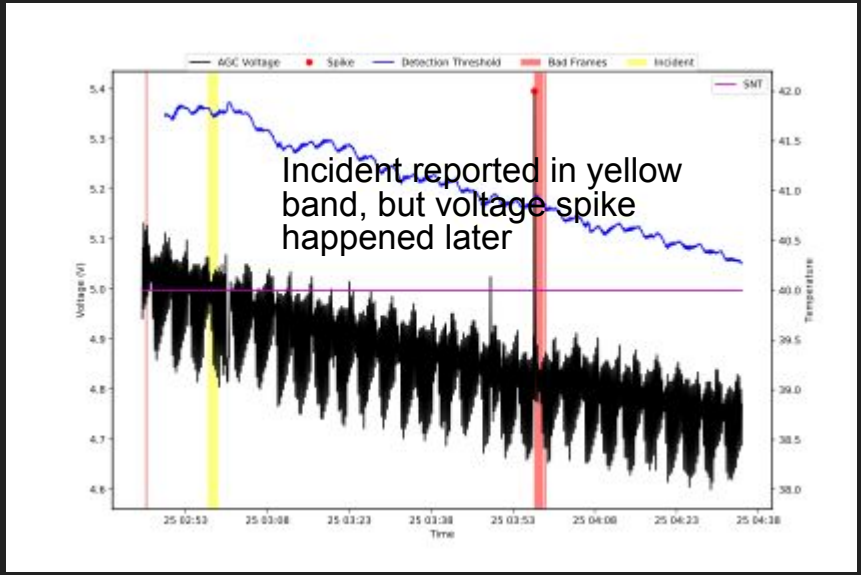
- Activity Comparisons
- Anomaly Detection
- Predicting DRs
- Profiling Early/Lateness



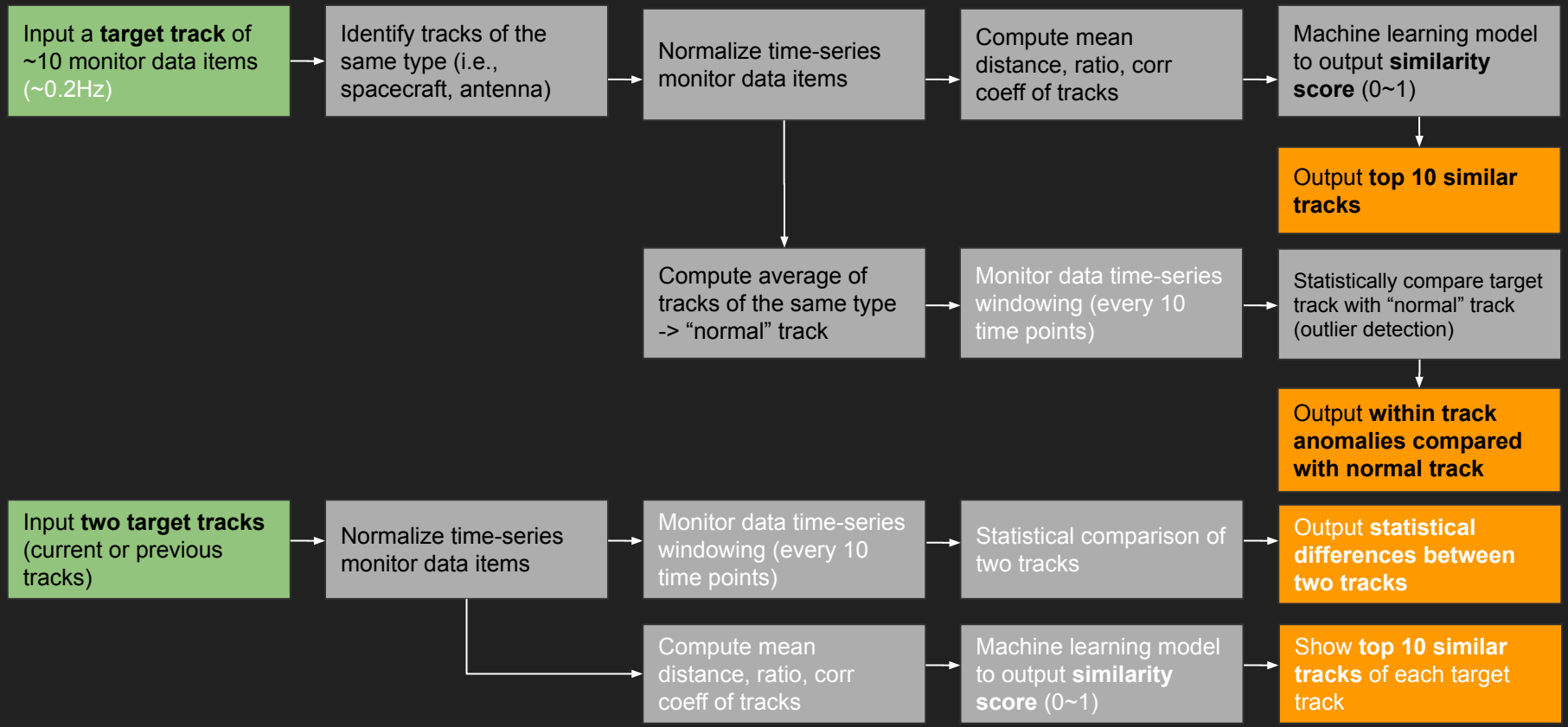
Classification based on Discrepancy Reports

- Top-ranked use case
- Most problematic due to issues with labeled data
- DSN to pursue a new tool that will improve label quality for ML



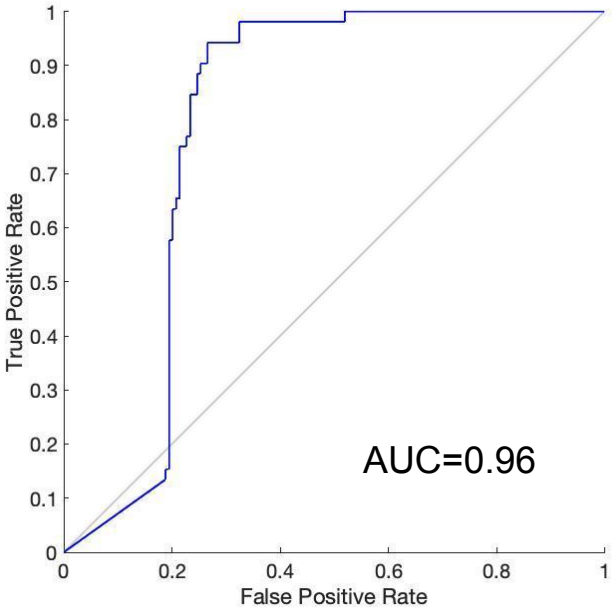


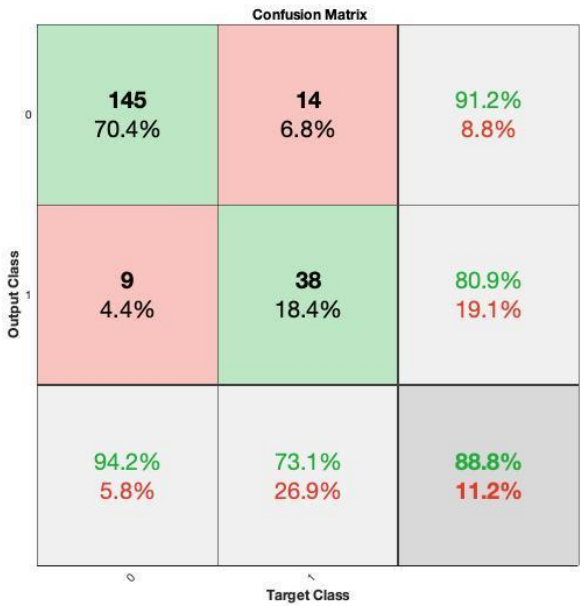
Track comparisons input-output architecture diagram



Track Comparisons - preliminary results (206 tracks: 52 similar & 154 dissimilar tracks)

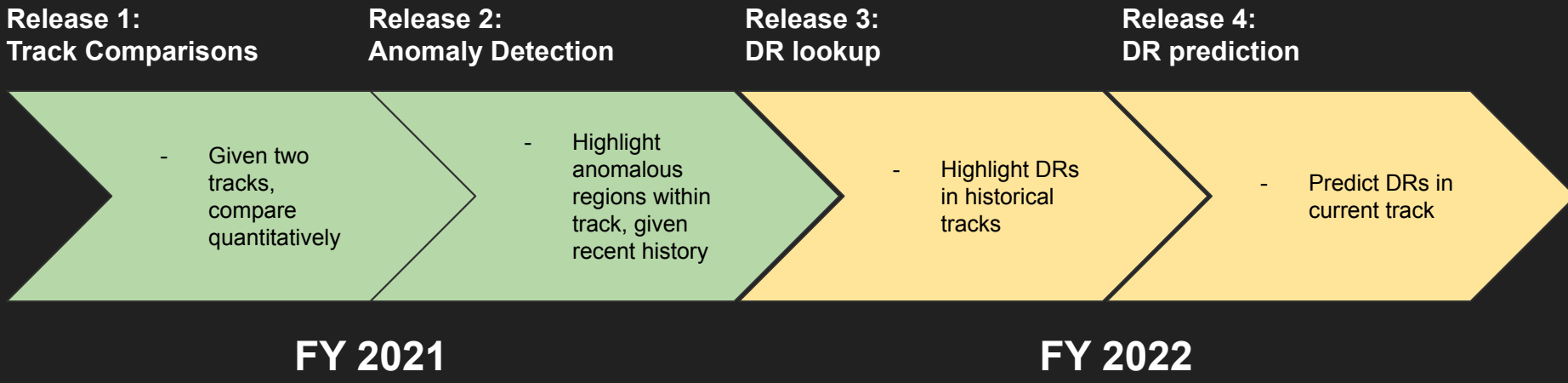
Model	Accuracy
Decision Tree	60%
Linear Regression	73.3%
Logistic Regression	73.3%
Naive Bayes	66.7%
Linear SVM	73.3%
KNN	73.3%
Ensemble Bagged Trees	80.0%
Neural Net	88.8%





Roadmap to Integration

The below roadmap details a development plan based on survey results on utility of tools, maturity of current tool development, and available resources.



Conclusion

- DSN is preparing for major shifts in its operational paradigm
- Machine learning and automation will certainly play a role
- Build trust and value for users
  - Survey needs
  - Integrate with simpler products
- Data quality must be addressed as we shift to supervised models